Compensating for missing plot observations in forest inventory estimation

Ronald E. McRoberts

Abstract: The Enhanced Forest Inventory and Analysis program of the U.S. Forest Service has established a nationwide array of permanent field plots, each representing approximately 2400 ha. Each plot has been assigned to one of five interpenetrating, nonoverlapping panels, with one panel selected for measurement on a rotating basis each year. As with most large surveys, observations for some plots are not obtained. Several techniques for dealing with missing observations were compared with respect to the bias and precision of stratified estimates of mean volume per unit area using data from the State of Indiana in the north-central region of the United States for illustrative purposes. For the proportion of missing plots ranging from 1% to 10%, acceptable results were obtained for techniques that both ignored and replaced missing plot observations, but only when the techniques accounted for differences between plots in private and public ownerships.

Résumé : Le programme amélioré d'inventaire forestier et d'analyse du Service forestier des États-Unis a permis d'établir un dispositif national de placettes permanentes. Représentant approximativement 2400 ha, chaque placette a été assignée à l'un des cinq groupes qui s'interpénètrent sans chevauchement. Chaque groupe est sélectionné à tour de rôle pour être mesuré sur une base annuelle. Comme dans tous les inventaires à grande échelle, les observations de certaines placettes manquent. Pour résoudre ce problème d'observations manquantes, plusieurs techniques ont été comparées quant au biais et à la précision des estimations stratifiées du volume moyen par unité de surface, en utilisant à titre d'exemple les données de l'État de l'Indiana situé dans le centre nord des États-Unis. Lorsque la proportion des placettes manquantes se situe entre 1 et 10 %, des résultats acceptables ont été obtenus tant par les techniques qui ignorent que celles qui remplacent les observations des placettes manquantes, mais seulement lorsque ces techniques tiennent compte des différences entre les placettes sur les terres publiques et celles sur les terres privées.

[Traduit par la Rédaction]

Introduction

The Forest Inventory and Analysis (FIA) program of the U.S. Forest Service, an agency of the U.S. Department of Agriculture, conducts inventories of U.S. forest land to determine its extent, condition, and volume of standing timber, timber growth, and timber removals. The FIA program has recently initiated a transition from periodic inventories in which states are selected on a rotating basis for multiyear complete inventories to annual inventories in which a proportion of plots is measured in each state each year (McRoberts 1999). The new system, designated the Enhanced FIA program, features an equal probability sample of permanent field plots, each representing approximately 2400 ha (Brand et al. 2000). In the eastern United States, this sample has been systematically divided into five interpenetrating, nonoverlapping panels. Each year, the plots in a single panel are selected for measurement, and panels are selected on a 5-year rotating basis. Inventory reports are required for each state every 5 years.

Received 25 September 2002. Accepted 11 April 2003. Published on the NRC Research Press Web site at http://cjfr.nrc.ca on 9 October 2003.

R.E. McRoberts. Forest Inventory and Analysis, North Central Research Station, USDA Forest Service, 1992 Folwell Avenue, St. Paul, MN 55108, U.S.A. (e-mail: rmcroberts@fs.fed.us).

Inevitably, observations for some plots are not obtained. Missing plot observations are attributed to a variety of causes: (*i*) plots are inaccessible because of hazardous terrain, (*ii*) plots are inaccessible because of environmental conditions (e.g., underwater in the year selected for measurement), (*iii*) plots are inaccessible because of cultural prohibitions (e.g., located in a Native American burial ground), (*iv*) landowners deny field crews access to plots, and (*v*) sufficient time to complete measurement of an entire plot is not available. Nationally, the proportion of missing plot observations varies by state and ranges from less than 1% to as great as 10%.

The topic of missing plot observations in forest inventory falls within the broader statistical context of nonresponse in sample surveys (Lemeshow 1985). A variety of techniques have been considered for compensating for nonresponses. Sometimes the missing observations are simply ignored and all analyses are performed on only the actual collected data. This approach is generally regarded as rather dangerous because of the possibly incorrect underlying assumption that the missing data are similar in all important regards to the collected data. As a result, imputation techniques that replace the missing observations with suitable values are usually considered. The objective of this study was to compare several techniques for dealing with missing observations or nonresponses in forest inventory with respect to the bias and precision of stratified estimates of mean volume per unit area obtained from forest inventory plot data. Data obtained for the State of Indiana in the north-central region of the

 doi: 10.1139/X03-112

United States were used to illustrate and compare the techniques.

Materials and methods

Data

For the FIA program of the North Central Research Station (NCRS), FIA field plot data include a variety of plot- and tree-level observations obtained in two phases. In phase 1, each plot is observed using aerial photography or digital orthoquads to determine if trees are growing on it. Plots determined in phase 1 to be without trees receive no additional observation, and values for all tree-related variables are set to zero. At the beginning of phase 2, FIA field crews consult county landowner records to determine plot ownership so that permission may be obtained to measure privately owned plots. They also assign each of these plots to an ownership category: federal, state, local public, corporate, or private. In phase 2, field crews visit field plots and obtain measurements of individual trees from which tree-level volumes (Hahn 1984; Smith 1985; Hahn and Hansen 1991) and plot-level volumes per unit area, V, are estimated.

Data were obtained for FIA plots measured in each of the two most recent periodic inventories of Indiana: 1986 (Smith and Golitz 1988) and 1998 (Schmidt et al. 2000). For the 1986 inventory, FIA plots consisted of 10-point clusters of variable-radius subplots, while for the 1998 inventory, plots consisted of clusters of four fixed-radius subplots. Although the plot configurations and sampling techniques for the two inventories differed, plot centers were coincident so that estimates of V pertain to the same geographic area. Data for all plots measured in both inventories, regardless of disturbance or other history, were included in this study.

Stratified estimation

Because of budgetary constraints and natural variability among plots, sample sizes sufficient to satisfy national FIA precision standards are seldom achieved for most inventory variables unless the estimation process is enhanced. Thus, ancillary data in the form of aerial photography (Hansen 1990; Loetsch and Haller 1964) or classified satellite imagery (Hansen and Wendt 2000; McRoberts et al. 2002) have been used to create strata to increase the precision of estimates with stratified estimation. The NCRS FIA program has used the National Land Cover Dataset (NLCD) (Vogelmann et al. 2001) as a basis for creating strata for stratified estimation. The NLCD is a 21-class, 30 m × 30 m pixel-based, land cover map of the conterminous United States based on satellite imagery and ancillary data. McRoberts et al. (2002) investigated the utility of the NLCD for stratification for variance reduction purposes and recommended creating four strata using a three-step process: (*i*) aggregate NLCD classes with trees into a forest stratum and the remaining classes into a nonforest stratum, (*ii*) reclassify isolated groups of three or fewer pixels into their surrounding class, and (*iii*) create a forest edge stratum by removing from the forest stratum a 2-pixel-wide band on the forest side of the forest–nonforest boundary and create a nonforest edge stratum by removing from the nonforest stratum a 2-pixel-wide band on the nonforest side of the forest–nonforest boundary. For stratified estimation, NCRS FIA uses these four strata: forest, forest edge, nonforest edge, and nonforest.

Stratification of a land area requires that each plot be assigned to a single stratum and that the relative proportion of the land area corresponding to each stratum be estimated. The first task is accomplished by assigning plots to strata on the basis of the strata assignments of their associated pixels. The second task is accomplished simply by counting the number of pixels with centers in each of the four strata and then calculating the relative proportions of pixels by strata. These proportions are used as strata weights when calculating stratified estimates of means and variances.

Stratified estimates of mean volume per unit area, $\bar{V}$, and variance, $\mathrm{Var}(\bar{V})$, were calculated using standard methods (Cochran 1977):

$$[1] \qquad \bar{V} = \sum_{j=1}^{J} w_j \bar{V}_j$$

and

$$[2] \qquad \mathrm{Var}(\bar{V}) = \sum_{j=1}^{J} w_j^2 \frac{\hat{\sigma}_j^2}{n_j}$$

where $j = 1,..., J$ denoted stratum, w_j was the weight for the jth stratum, calculated as the proportion of pixels assigned to the stratum, $\bar{V}_j$ was the mean volume per unit area for plots assigned to the jth stratum, n_j was the number of plots assigned to the jth stratum, and $\hat{\sigma}_j^2$ was the within-stratum variance for the jth stratum calculated as

$$[3] \qquad \hat{\sigma}_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (V_{ji} - \bar{V}_j)^2$$

where V_{ji} was the volume per unit area observed by the field crew for the ith plot assigned to the jth stratum. Variance estimates obtained using eq. 2 ignored the slight effects due to finite population correction factors and to variable rather than fixed numbers of plots per strata.

Approaches to compensating for missing observations

Approaches to compensating for missing plot observations in inventory estimation included ignoring the fact that observations for some plots are missing or replacing the missing observations. Although all techniques that replace missing observations could be, and frequently are, characterized as imputation techniques, for purposes of clarity in the discussions and analyses that follow, four categories of these techniques were distinguished with respect to the nature of the replacements: previous observations, stratum means for observed plots, model predictions, and nearest-neighbor imputations from observed plots. In the north-central region of the United States, virtually all missing plot observations may be attributed to private landowners who deny field crews access to plot locations. Thus, for this study, all plots with missing observations were assumed to be privately owned.

Standard for comparison

Stratified estimates of $\bar{V}$ and $\mathrm{Var}(\bar{V})$ were calculated using observations for all plots with the four NLCD-derived strata. These estimates, designated COMPARE, served as the standard for comparison for estimates obtained with the tech-

Table 1. Stratum statistics.

					Volume per unit area (m^3/ha)					
					1986		1998			
	Weight		Number of plots		Mean		Mean		SE	
Stratum and substratum	Stratum	Substratum	Stratum	Substratum	Stratum	Substratum	Stratum	Substratum	Stratum	Substratum
Forest	0.119		803		81.31		106.77		2.50	
Nontree		0.012		81		0.00		0.00		0.00
Private		0.090		606		92.86		123.66		11.47
Public		0.017		116		42.70		75.96		6.10
Forest edge	0.107		760		56.90		76.68		2.76	
Nontree		0.029		207		0.00		0.00		0.00
Private		0.058		410		75.42		95.44		3.55
Public		0.020		143		86.22		124.73		5.86
Nonforest edge	0.150		916		10.21		16.20		1.31	
Nontree		0.116		713		0.00		0.00		0.00
Private		0.030		181		47.10		58.27		4.60
Public		0.004		22		37.68		50.74		13.86
Nonforest	0.625		3847		2.05		2.59		0.31	
Nontree		0.601		3700		0.00		0.00		0.00
Private		0.019		117		45.70		47.94		6.38
Public		0.005		30		84.95		123.66		11.47
Statewide	1.000		6326		18.55		24.92		0.50	

niques that compensate for missing plot observations. COMPARE estimates were also calculated using substrata obtained by subdividing each of the four original strata with respect to factors that affect V. The rationale for subdividing strata and the nature of the subdivisions are discussed in the sections that follow.

Ignoring plots with missing observations

Two techniques were investigated that ignored the fact that some plots have missing observations. For both, the area represented by each observed plot within strata increased, while the precision of strata means and the overall mean decreased because of reduced sample sizes. The first technique, designated IGNORE, treated the plots with missing observations as if they had not been selected for the sample and calculated estimates in the usual manner, albeit with slightly smaller sample sizes.

If missing plot observations are not missing at random, then ignoring the fact that they are missing may result in biased estimates. In many areas of the country, privately owned forest lands are managed differently than forest lands in public ownership, with the result that V may differ with respect to ownership (Table 1). For example, if V is generally greater on privately owned forest lands, then ignoring the fact that observations for some privately owned plots are missing causes plots in public ownership with smaller V to be overrepresented in strata means. One approach is to subdivide strata with missing plot observations with respect to ownership.

The second technique, designated IGNORE-S, was similar to IGNORE except that each stratum with missing plot observations was subdivided into two or three substrata depending on available information. To avoid confusion between the four original strata and the subdivisions of those strata, the strata subdivisions are referred to as substrata, although they were treated as separate strata when calculating stratified estimates. If a digital landownership layer that can be used with a geographic information system (GIS) is available, then subdivision of each stratum into private and public substrata with GIS techniques suffices. Plots are assigned to substrata on the basis of the substrata assignments of their associated pixels. When using stratified estimation, the method for calculating substrata weights must be consistent with the method for assigning plots to substrata. Thus, if a digital ownership layer is used to calculate substrata weights, then that same layer must be used to assign plots to substrata, even if assignments disagree with the county landowner records consulted by the field crew. Pixels assigned to each of the four original strata are further assigned to private or public substrata using the digital ownership layer with a GIS, and substrata weights are calculated in the same manner as the four original strata weights are calculated.

If a digital landownership layer is unavailable, then an ad hoc procedure is necessary to subdivide each stratum into three substrata: nontree, private, and public. The first task, assigning plots to substrata, will still be easy. Plots determined in phase 1 to be without trees are assigned to nontree substrata because ownership of these plots is generally not determined; plots determined in phase 1 to have trees are assigned to private or public substrata on the basis of county landowner records. The second task, determining the relative proportions of strata that are in nontree, private, and public substrata, is more difficult. However, because the plots are an equal probability sample of the land area of interest, within-strata ratios of the number of nontree, privately owned, and publicly owned plots to the total number of plots are unbiased estimates of the relative proportions of strata that are without trees, privately owned, and publicly owned, respectively. Thus, weights for nontree substrata, for example, are calculated as products of the original strata weights obtained from the phase 1 pixel counts and the proportions of plots in nontree substrata; weights for the private and public substrata are calculated in a similar manner. Stratified estimates are then based on the 12 substrata rather than on the original four strata. For this study, a digital landownership layer was assumed to be unavailable, and the ad hoc stratification procedure was used.

Replacement with previous observations

All replacement techniques were based on three assumptions: (*i*) plots with missing observations for two consecutive inventories were either replaced at the time of the second inventory or were removed from the sample, (*ii*) the plot array remained unchanged over time, with the exception of plots replaced because of missing observations for consecutive inventories, and (*iii*) all plots were remeasured at a constant or nearly constant time interval. The first assumption assured that plots with missing observations in the current inventory were measured in the previous inventory, and the second and third assumptions assured the availability of observed plots that were similar to plots with missing observations.

The third technique, designated PREVIOUS, replaced each missing plot observation with the observation for the same plot from the previous inventory. The rationale for this technique was that V for the current inventory could be expressed as the sum of V for the previous inventory and the change in V between inventories. For most plots, V for the previous inventory was a much larger proportion of current V than was change in V, particularly for the relatively short 5-year Enhanced FIA remeasurement interval. Thus, the PREVIOUS technique captured the greater proportion of current V and was easy to implement. However, the remeasurement interval between the previous and current inventories for data used in this study was 12 years, considerably greater than the 5-year interval for the Enhanced FIA program. Thus, for the PREVIOUS technique, V for the previous inventory was adjusted to represent V 5 years before the current inventory under the assumption of constant annual change in V between inventories.

Replacement with strata means

A missing plot observation may be replaced with the mean for observed plots for the same stratum to which a plot with a missing observation is assigned. However, because strata means may be biased when all missing observations are for privately owned plots, strata with missing plot observations were subdivided in the same manner as for the IGNORE-S technique. The fourth technique, designated STRATUM, replaced each missing observation with the mean for the private substratum of the stratum to which the plot was originally assigned in phase 1.

Although the STRATUM technique yields an unbiased estimate of the substratum mean, the substratum variance will be biased downward because observations for all missing plot observations are replaced with the same value. To compensate, the fifth technique replaced each missing plot observation with the sum of the substratum mean and a number, u, representing uncertainty from two sources, the uncertainty in the substratum sample mean and observed variability around this mean. For each missing observation, u was independently and randomly generated from a $N(0, \sigma)$ distribution with σ incorporating the variability from both sources. This technique was designated STRATUM+U.

The process of randomly selecting and incorporating u may be considered an imputation. Rubin (1987) advocated multiple completions of data sets via imputation to allow assessment of the uncertainty in imputed variables and to protect against extreme results. Thus, for the STRATUM+U technique, each set of missing observations was replaced independently m times, m separate stratified estimates of the mean and variance were calculated, and the m separate estimates were combined as recommended by Rubin (1987):

$$[4] \qquad \bar{V} = \frac{1}{m}\sum_{k=1}^{m} \bar{V}^k$$

and

$$[5] \qquad \mathrm{Var}(\bar{V}) = \frac{1}{m}\sum_{k=1}^{m} \mathrm{Var}(\bar{V}^k) + \frac{m+1}{m}\sigma^2_{\bar{V}}$$

where $\bar{V}^k$ and $\mathrm{Var}(\bar{V}^k)$ were the stratified estimates of the mean and variance for the kth completion of the data set and $\sigma^2_{\bar{V}}$ was the variance among the m stratified estimates of the mean. For this study, $m = 5$, which was even larger than the $m = 2$ or $m = 3$ found to be adequate in unrelated studies by Rubin and Schenker (1986) and Reams and McCollum (2000).

If V differs substantially with respect to ownership, then stratified estimates based on the substrata may contribute to increasing the precision of estimates beyond the increase in precision obtained using the four original strata alone. For both the STRATUM and STRATUM+U techniques, stratified estimates were calculated using the 12 substrata. The stratified estimates obtained using the 12 substrata obtained with the ad hoc stratification procedure were assumed to be comparable with estimates that would have been obtained using the digital landownership layer. Because substrata statistics were calculated to implement the STRATUM and STRATUM+U techniques, little additional effort was necessary to calculate stratified estimates using the 12 substrata.

Replacement with model predictions

Two model-based techniques for replacing missing plot observations were investigated. Both were based on simple linear regression models of the relationships between current and previous V for observed plots in private substrata. The first model-based technique and the sixth technique overall, designated MODEL, replaced each missing observation with the predicted V obtained from the model for the appropriate private substratum. However, the MODEL technique used the same model prediction to replace all missing observations with the same previous V and therefore, as with the STRATUM technique, may produce downward-biased estimates of substrata variances. The second model-based technique and the seventh overall technique replaced each missing observation with the sum of the model prediction and a number, u, representing uncertainty from two sources, the uncertainty in the prediction resulting from the sample-based estimates of the model parameters and observed residual variability around this prediction. For each missing observation, u was independently and randomly generated from a $N(0, \sigma)$ distribution with σ incorporating the variability from both sources. This technique was designated MODEL+U, and because of its imputation nature, each set of missing observations was replaced independently five times, five separate stratified estimates of the mean and variance were calculated, and the five separate estimates were combined using eqs. 4 and 5. Stratified estimates for both the MODEL and MODEL+U techniques were also calculated using the 12 substrata.

An advantage of model-based techniques is that data for calibrating the substrata models need not be restricted to plots observed in the immediately previous and current inventories. If relationships are temporally and regionally stable, then data from inventories before the immediately previous and current inventories and (or) data from other geographic regions may be used to calibrate the models. However, adjustments may be necessary to compensate for differences in time intervals between inventories. For this study, only observations for the immediately previous and current inventories were used to calibrate the models.

Replacement with imputations

Imputation-based techniques select replacements for missing plot observations from current observations. Sande (1982) distinguished between hotdeck procedures that select from a pool of current observations that exactly match the missing observations with respect to prescribed attributes and nearest-neighbor procedures that select from a pool of current observations that are only similar to the missing observations with respect to the prescribed attributes. Van Deusen (1997) and Reams and Van Deusen (1999) proposed using the imputation techniques discussed by Sande (1982, 1983), Rubin and Schenker (1986), and Rubin (1987) for replacing missing inventory observations. Reams and McCollum (2000) reported on hotdeck imputation procedures for updating estimates of V for plots measured in previous years.

Two nearest-neighbor imputation techniques were investigated. For both techniques, a pool of the five observed plots most similar to the plot with the missing observation with respect to previous V was selected, one of the five plots was randomly selected, and V for the selected plot was imputed to the plot with the missing observation. The first imputation technique and the eighth technique overall, designated IMPUTE, selected the pool of five most similar plots from among all privately owned, observed plots without regard to their strata assignments. The second imputation technique and the ninth technique overall, designated IMPUTE-S, was similar to the IMPUTE technique except the pool of five most similar, privately owned, observed plots was further restricted to the same stratum as the plot with the missing observation. For both imputation techniques, each set of missing observations was replaced independently five times, five separate stratified estimates of the mean and variance were calculated, and the five separate estimates were combined using eqs. 4 and 5. Stratified estimates for the IMPUTE-S technique were also calculated using the 12 substrata.

Comparisons

Estimates for the nine techniques were compared with the COMPARE estimates and with each other for proportions of missing plot observations ranging from 1% to 10%. The proportion was calculated as the ratio of the number of plots classified as denied access to the total number of plots selected for measurement by field crews. The denominator of this ratio did not include plots determined in phase 1 to be without trees because field crews make no attempt to visit these plots.

Simulations were used to mimic random private landowner denials of field crew access to plots and to implement the techniques for dealing with the missing observations that required random selections: STRATUM+U, MODEL+U, IMPUTE, and IMPUTE-S. For these analyses, private landowner denials were assumed to be a random phenomenon. The simulation procedure generated distributions of $\bar{V}$ and Var($\bar{V}$) for randomly selected combinations of missing plot observations using a four-step procedure as follows.

(1) Initialization:

(*a*) Calculate weights for the four original strata.

(*b*) Calculate weights for nontree, private, and public substrata based on the four original strata and proportions of nontree, privately owned, and publicly owned plots within strata.

(*c*) Calculate COMPARE stratified estimates of $\bar{V}$ and Var($\bar{V}$) using observations for all plots.

(2) Plots with missing observations: Randomly select from all privately owned plots, the proportion designated as having missed observations.

(3) Techniques for dealing with missing plot observations:

(*a*) Calculate stratified estimates of $\bar{V}$ and Var($\bar{V}$) using the IGNORE, IGNORE-S, and PREVIOUS techniques.

(*b*) Calculate substrata means and variances. (*i*) Calculate STRATUM replacements and calculate stratified estimates of $\bar{V}$ and Var($\bar{V}$). (*ii*) Calculate five independent sets of STRATUM+U replacements, calculate five stratified estimates of $\bar{V}$ and Var($\bar{V}$), and combine the estimates using eqs. 4 and 5.

(*c*) Estimate the coefficients of substrata models of the relationships between current and previous V. (*i*) Calculate MODEL replacements and calculate stratified estimates of $\bar{V}$ and Var($\bar{V}$). (*ii*) Calculate five independent sets of MODEL+U replacements, calculate five sets of stratified estimates of $\bar{V}$ and Var($\bar{V}$), and combine the estimates using eqs. 4 and 5.

(*d*) Select five independent sets of IMPUTE and IMPUTE-S replacements, calculate five sets of stratified estimates of $\bar{V}$ and Var($\bar{V}$), and combine the estimates using eqs. 4 and 5.

(*e*) Retain stratified estimates of $\bar{V}$ and Var($\bar{V}$) for each technique.

(4) Repetition: Repeat steps 2 and 3 1000 times.

For each designated proportion of missing plot observations and for each of the nine techniques, the means of the distributions of the 1000 simulated stratified estimates of $\bar{V}$ and Var($\bar{V}$) were calculated. The bias and precision of estimates associated for each technique were compared using these means.

Results

The means of the distributions of stratified estimates of $\bar{V}$ stabilized by 1000 simulations for all techniques (Fig. 1), indicating that 1000 simulations was adequate.

Table 2. Stratified estimates of mean volume per unit area (m^3/ha) and standard error of the mean.

	Proportion of plots missing											
	0.00			0.01			0.05			0.10		
		SE			SE			SE			SE	
Technique	Mean	4 strata	12 strata	Mean	4 strata	12 strata	Mean	4 strata	12 strata	Mean	4 strata	12 strata
COMPARE	24.92	0.50	0.41									
IGNORE				24.89	0.50		24.77	0.50		24.61	0.51	
IGNORE-S				24.92		0.41	24.92		0.41	24.93		0.41
PREVIOUS				24.91	0.50		24.87	0.50		24.82	0.50	
STRATUM				24.92	0.50	0.41	24.92	0.50	0.41	24.93	0.50	0.40
STRATUM+U				24.92	0.50	0.41	24.91	0.50	0.41	24.90	0.51	0.41
MODEL				24.92	0.50	0.41	24.92	0.50	0.41	24.93	0.49	0.41
MODEL+U				24.92	0.50	0.41	24.94	0.50	0.41	24.96	0.50	0.41
IMPUTE				24.91	0.50		24.88	0.50		24.85	0.51	
IMPUTE-S				24.92	0.50	0.41	24.92	0.50	0.42	24.93	0.51	0.42

Fig. 1. Means of distributions of STRATUM+U stratified estimates of mean volume per unit area by simulation.

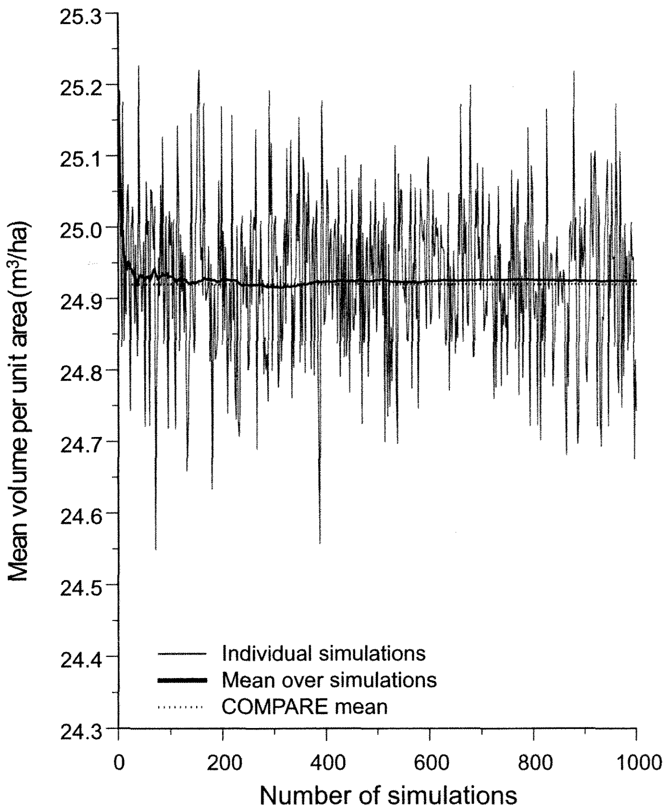
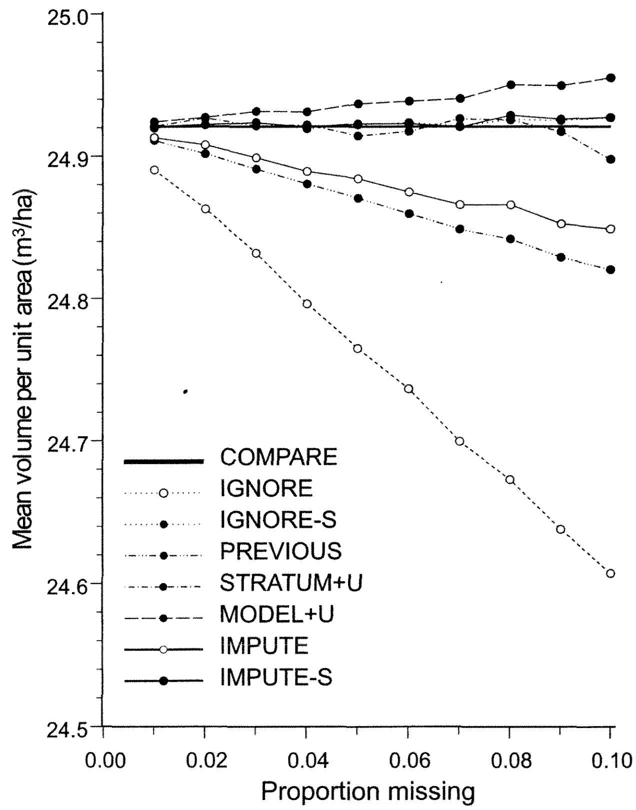


Fig. 2. Means of distributions of stratified estimates of mean volume per unit area.



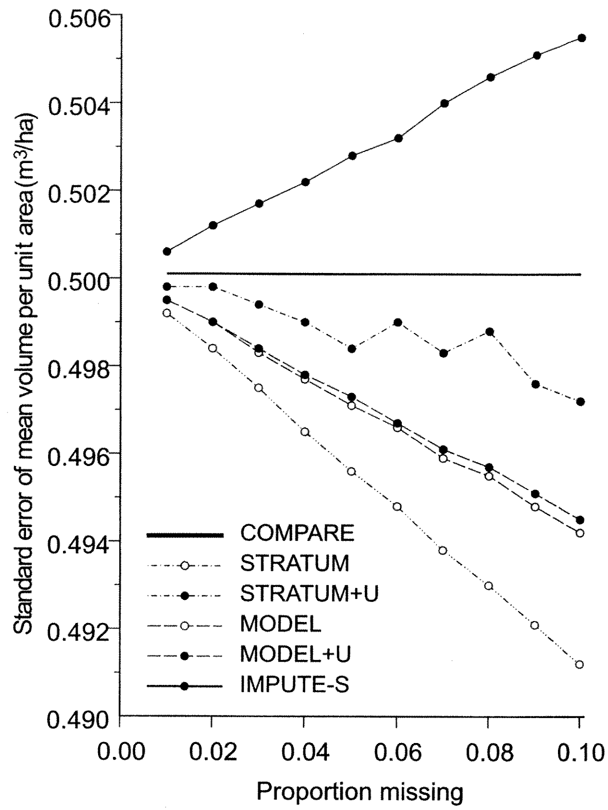
An important result was that the means of the distributions of simulated stratified estimates of $\bar{V}$ obtained with the nine techniques for dealing with missing observations were all well within one standard error of the COMPARE mean (Table 2). This result indicated there were no statistically significant differences between any of the nine stratified estimates of $\bar{V}$ and the COMPARE estimate. Nevertheless, there were several discernible trends among the estimates for the various techniques (Fig. 2). First, means for the IGNORE, IMPUTE, PREVIOUS, MODEL, and MODEL+U techniques increasingly diverged from the COMPARE estimate as the proportion of missing plot observations increased, while the means for the remaining techniques followed closely the COMPARE estimate. Failure of the means for the IGNORE and IMPUTE techniques to follow closely the COMPARE estimate compared with the IGNORE-S and IMPUTE-S means indicated that failure to account for differences in V with respect to ownership adversely affected the quality of estimates. Second, failure of means for the PREVIOUS technique to follow closely the COMPARE estimate indicated that failure to account for change in V since the previous inventory also adversely affected the quality of estimates.

Third, failure of both the MODEL and MODEL+U means to follow closely the COMPARE estimate, even though both account for ownership differences, indicated that a simple linear model for the relationship between V and previous V was probably not adequate. Finally, there was little to distinguish the IGNORE-S, STRATUM+U, and IMPUTE-S estimates of $\bar{V}$ from each other or from the COMPARE estimate.

The standard error of $\bar{V}$ was estimated for each of the nine techniques as the square root of mean Var($\bar{V}$) for the 1000 simulations. Stratified estimates of standard errors based on the four original strata were of the same order of magnitude as the COMPARE estimate based on the same strata, and estimates based on the 12 substrata were of the same order of magnitude as the COMPARE estimate based on the 12 substrata (Table 2). However, the 12 substrata estimates were typically 20% to 25% less than estimates corresponding to the four original strata.

Several general trends were evident when comparing estimates of standard errors by technique. First, inclusion of randomly generated residuals when calculating replacement values using the STRATUM+U and MODEL+U techniques partially compensated for the downward bias in the estimates obtained using the STRATUM and MODEL techniques (Fig. 3). However, compensation by the MODEL+U technique was barely discernable, while compensation by the STRATUM+U technique, while relatively substantial, still produced estimates that underestimated the COMPARE estimate. Second, between-completions variability when using eqs. 4 and 5 to combine estimates for the five completions of data sets was small. Third, stratified estimates of standard errors based on the original four strata for the STRATUM+U technique more closely followed the COMPARE standard error estimate than did estimates for the other techniques with estimates of $\bar{V}$ that closely followed the COMPARE mean estimate. Fourth, deviations from the COMPARE standard error increased for all techniques as the proportion of missing plot observations increased, although none deviated by more than 2%. Finally, the techniques were similar with respect to estimates of standard errors based on the original four strata and also similar with respect to estimates based on the 12 substrata.

Fig. 3. Means of distributions of stratified estimates of the standard error of mean volume per unit area.



Conclusions

Several conclusions were drawn from this study. First, the factor with the greatest effect on standard errors was variability among plots. The effects of this factor were to render as statistically nonsignificant differences between the COMPARE stratified mean estimates and mean estimates for all nine of the techniques investigated for compensating for missing plot observations. In addition, the variability among plots was much greater than the between-completions variability, thus rendering as small in a relative sense the effect of the latter on standard errors. Second, consideration should be given to factors such as ownership, particularly if there is reason to think that management of forest lands or growth and survival characteristics differ by the categories of these factors. Ignoring the fact that some observations were missing produced increasingly large deviations in estimates as the proportion of missing plots increased unless the stratified estimation used substrata that separated the plots with respect to categories of ownership. Similarly, replacement techniques produced increasingly large deviations in estimates as the proportion of missing plots increased unless the replacements were based on observed plots in the privately owned substrata only. Third, separation of substrata based on categories of such factors may increase the precision of stratified estimates. Separation of nontree, privately owned, and publicly owned substrata increased the precision of estimates by 20% to 25%, although this finding should be confirmed using digital ownership layers for obtaining substrata weights.

Finally, results from this study lead to a two-part recommendation for dealing with missing plot observations. First, if users are comfortable with ad hoc procedures for obtaining substrata weights or if digital ownership layers are available, then the IGNORE-S technique is appropriate. It was the easiest to implement, and it produced estimates that closely followed the COMPARE estimates and captured the increase in precision that accrues from using more strata. Second, if users are uncomfortable with ad hoc procedures and an ownership layer is not available, then the STRATUM+U technique with the original strata is an appropriate selection. This technique produced estimates that closely followed the COMPARE estimate and that were nearly indistinguishable from the IGNORE-S, STRATUM, and IMPUTE-S estimates. Further, the STRATUM+U technique produced estimates of standard errors that were closest to the COMPARE estimate, and although the estimates were less than the COMPARE estimate, the maximum deviation was negligible. Finally, the STRATUM+U technique was easier to implement than the MODEL, MODEL+U, and IMPUTE techniques, did not require ad hoc procedures, and was only slightly more difficult to implement than the STRATUM technique.

References

Brand, G.J., Nelson, M.D., Wendt, D.G., and Nimerfro, K.N. 2000. The hexagon panel system for selecting FIA plots for an annual inventory. *In* Proceedings of the First Annual Forest Inventory and Analysis Symposium, 2–3 November 1999, San Antonio, Tex. *Edited by* R.E. McRoberts, G.A. Reams, and P.C. Van Deusen. U.S. For. Serv. Gen. Tech. Rep. NC-213. pp. 8–13.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. Wiley, New York.

Hahn, J.T. 1984. Tree volume and biomass equations for the Lake States. U.S. For. Serv. Res. Pap. NC-250.

Hahn, J.T., and Hansen, M.H. 1991. Cubic and board foot volume models for the Central States. North. J. Appl. For. **8**: 47–57.

Hansen, M.H. 1990. A comprehensive sampling system for forest inventory based on an individual tree growth model. Ph.D. dissertation, College of Natural Resources, University of Minnesota, St. Paul, Minn.

Hansen, M.H., and Wendt, D.G. 2000. Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. *In* Proceedings of the First Annual Forest Inventory and Analysis Symposium, 2–3 November 1999, San Antonio, Tex. *Edited by* R.E. McRoberts, G.A. Reams, and P.C. Van Deusen. U.S. For. Serv. Gen. Tech. Rep. NC-213. pp. 20–27.

Lemeshow, S. 1985. Nonresponse (in sample surveys). *In* Encyclopedia of statistics. Vol. 6. *Edited by* S. Kotz and N.L. Johnson. John Wiley & Sons, New York. pp. 333–336.

Loetsch, F., and Haller, K.E. 1964. Forest inventory. Vol. I. Statistics of forest inventory and information from aerial photographs. BLV Verlagsgesselschaft, Munich.

McRoberts, R.E. 1999. Joint annual forest inventory and monitoring system: the North Central perspective. J. For. **97**(12): 27–31.

McRoberts, R.E., Wendt, D.G., Nelson, M.D., and Hansen, M.H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sens. Environ. **81**: 36–44.

Reams, G.A., and McCollum, J.M. 2000. The use of multiple imputation in the southern annual forest inventory system. *In* Integrated tools for natural resources inventories. Proceedings of the IUFRO Conference, 16–20 August 1998, Boise, Idaho. *Edited by* M.H. Hansen and T. Burk. U.S. For. Serv. Gen. Tech. Rep. NC-212. pp. 228–233.

Reams, G.A., and Van Deusen, P.C. 1999. The southern annual forest inventory system. J. Agric. Biol. Environ. Stat. **4**: 346–360.

Rubin, D.B. 1987. Multiple imputation in non-response surveys. Wiley, New York.

Rubin, D.B., and Schenker, N. 1986. Multiple imputation for interval estimation for simple random samples with ignorable nonresponse. J. Am. Stat. Assoc. **81**: 366–374.

Sande, I.G. 1982. Imputation in surveys: coping with reality. Am. Stat. **36**: 145–152.

Sande, I.G. 1983. Hot-deck imputation procedures. *In* Incomplete data in sample surveys. *Edited by* W.G. Madow and I. Olkin. Academic Press, New York. pp. 334–350.

Schmidt, T.L., Hansen, M.H., and Solomakos, J.A. 2000. Indiana's forests in 1998. U.S. For. Serv. Resour. Bull. NC-196.

Smith, W.B. 1985. Factors and equations to estimate forest biomass in the North Central Region. U.S. For. Serv. Res. Pap. NC-268.

Smith, W.B., and Golitz, M.F. 1988. Indiana's forest statistics, 1986. U.S. For. Serv. Resour. Bull. NC-180.

Van Deusen, P.C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Can. J. For. Res. **27**: 379–384.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, N. 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogram. Eng. Remote Sens. **67**: 650–662.